Acidianus filamentous virus 1 coat proteins display a helical fold spanning the filamentous archaeal viruses lineage

Adeline Goulet[a], Stéphanie Blangy[a], Peter Redder[b], David Prangishvili[b], Catarina Felisberto-Rodrigues[a], Patrick Forterre[b,c], Valérie Campanacci[a], and Christian Cambillau[a,1]

[a]Architecture et Fonction des Macromolécules Biologiques, Centre national de la recherche scientifique and Universités Aix-Marseille I & II, Architecture et Fonction des Macromolécules Biologiques, Unité Mixte de Recherche 6098, Case 932, 163 avenue de Luminy, 13288 Marseille Cedex 9, France; [b]Institut Pasteur, Unité de Biologie Moléculaire du Gène chez les Extrêmophiles, 28 Rue du Dr Roux, 75724 Paris Cedex 15, France; and [c]Institut de Génétique et Microbiologie, Université Paris-Sud and Centre national de la recherche scientifique, Unité Mixte de Recherche 8621, 91405 Orsay Cedex, France

Edited by Michael G. Rossmann, Purdue University, West Lafayette, IN, and approved October 19, 2009 (received for review September 1, 2009)

***Acidianus* filamentous virus 1 (AFV1), a member of the *Lipothrixviridae* family, infects the hyperthermophilic, acidophilic crenarchaeaon *Acidianus hospitalis.* The virion, covered with a lipidic outer shell, is 9,100-Å long and contains a 20.8-kb linear dsDNA genome. We have identified the two major coat proteins of the virion (MCPs; 132 and 140 amino acids). They bind DNA and form filaments when incubated with linear dsDNA. A C-terminal domain is identified in their crystal structure with a four-helix-bundle fold. In the topological model of the virion filament core, the genomic dsDNA superhelix wraps around the AFV1–132 basic protein, and the AFV1–140 basic N terminus binds genomic DNA, while its lipophilic C-terminal domain is imbedded in the lipidic outer shell. The four-helix bundle fold of the MCPs from AFV1 is identical to that of the coat protein (CP) of *Sulfolobus islandicus* rod-shaped virus (SIRV), a member of the *Rudiviridae* family. Despite low sequence identity between these proteins, their high degree of structural similarity suggests that they could have derived from a common ancestor and could thus define an yet undescribed viral lineage.**

crystal structure | viral coat proteins | viral lineages | virus evolution

Viruses are elusive living entities that produce virions to disseminate their genes. Although the origin and nature of viruses remain controversial (1–5), structural studies have started to shed light on the origin and evolution of virion components (6). Several structures of viral major coat proteins (MCPs) have been reported, for instance, the four helical bundle of the ssRNA Tobamovirus tobacco mosaic virus (TMV) (7), the α/β fold conserved among the small RNA bacteriophages belonging to the Leviviridae family (8), and the similar three-domain glycoproteins from alphaviruses and flaviviruses, suggesting that these virion components have a common origin (9, 10). Moreover, MCPs and capsid structures from dsDNA icosahedral viruses have revealed evolutionary relationships spanning different domains of life and led to the definition of two major ancient lineages of DNA viruses (11). The first one is characterized by the jelly-roll fold, single or double (12–14), first detected in the capsid of adenovirus and bacteriophage PRD1 (15) and later on in large eukaryotic dsDNA viruses (NCLDV) and in the archaeal virus STIV (12–14). The second one, exemplified by the MCP of bacteriophage HK97, has been detected in bacterial head-and-tail viruses, in herpes viruses and in archaeal and bacterial encapsulin (icosahedral intracellular vesicles of possible viral origin) (11, 16). All these data indicate that virions, as a specific mode of virus dissemination, originated several times independently in early life evolution. No doubt that many other transdomain lineages remain to be discovered that may encompass most of viral families.

A major goal in viral studies is now to determine how many viral lineages can be identified from structural studies of proteins involved in virion structure. Because most viruses infecting hyperthermophilic archaea produce virions with unique morphotypes, they are a priori good starting points for the discovery of protein families and possible identification of additional viral lineages (5, 17). It is in particular important to determine if the MCPs present in the linear virions of some archaeal viruses correspond to undescribed families. These viruses, which infect hyperthermophilic crenarchaea living in acidic hot springs (temperature >80 °C and pH <3), are radically different in their properties from viruses that infect bacteria and eukarya. They have dsDNA genomes, whereas bacterial and eukaryotic viruses with linear virions all have either ssDNA genomes or RNA genomes. Furthermore, ≈70–90% of their putative genes do not share any sequence homology with those of other viruses or cellular life forms, apart from other archaeal viruses of the same families (17).

Archaeal viruses producing linear virions have been classified into two families, the stiff rod-shaped Rudiviridae and the flexible filamentous Lipothrixviridae. In contrast to members of the other archaeal viral families for which genome sequences are available, several orthologous genes, including some encoding glycosyl transferases and transcriptional regulators, are shared by the rudiviruses and lipothrixviruses (17). Nine genes, among the putative 45 of the rudivirus *Sulfolobus islandicus* rod-shaped virus 1 (SIRV1), share orthologues with the lipothrixvirus *Sulfolobus islandicus* filamentous virus (SIFV). Based on these observations, a common ancestor to these linear archaeal viruses was suggested, and both families were classified into a different viral order, the "*Ligamenvirales*" (18).

Rudiviruses are stiff rod-shaped particles without envelope and vary considerably in length (average size 23 × 610–900 nm) proportionally to the size of the linear dsDNA genome for the four characterized rudiviruses SIRV1–2, *Acidianus* rod-shaped virus 1 (ARV1) and *Stygiolobus* rod-shaped virus, SRV (19–21). The virion is a tube-like superhelix constituted by the genome and copies of a single, glycosylated, basic DNA-binding protein (ORF134). In optimal conditions of temperature and pH (75 °C, pH3) and without additional energy source, the coat protein (CP) alone can self-assemble to produce filamentous structures

Author contributions: C.C. designed research; A.G., S.B., P.R., C.F.-R., and V.C. performed research; A.G., P.R., D.P., P.F., V.C., and C.C. analyzed data; and A.G., D.P., P.F., and C.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in Protein Data Bank, www.pdb.org (PDB ID codes 3FBZ (AFV1–140) and 3FBL (AFV1–132).

[1]To whom correspondence should be addressed. E-mail: christian.cambillau@afmb.univ-mrs.fr.

This article contains supporting information online at www.pnas.org/cgi/content/full/0909893106/DCSupplemental.

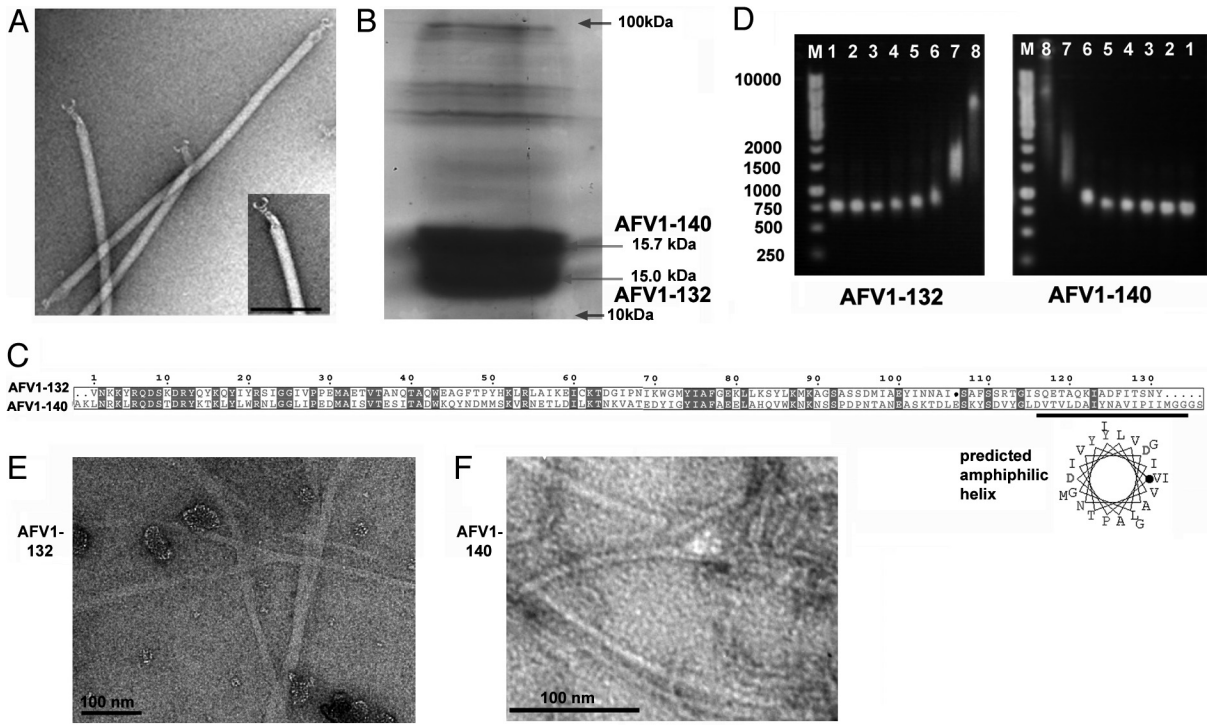

Fig. 1. The major coat proteins of AFV1. (*A*) Electron micrograph of AFV1 particles with tail structures in their native conformation [Bars, 100 nm; (22)]. (*B*) SDS gel (15% polyacrylamide) of AFV1 virion. Mass analysis indicates that only the two major bands belong to proteins from virions, namely AFV1–132 and AFV1–140, while the minor bands belong to the host. (*C*) Sequence alignment of AFV1–140 and AFV1–132 (alignment made by ClustalW, www.ebi.ac.uk/clustalw/, and displayed with ESPRIPT 2.2, www.espript.ibcp.fr/). The helical wheel diagram illustrates the amphiphilic character of the AFV1–140 C-terminal helix (residues 116–138, http://bp.nuap.nagoya-u.ac.jp/sosui/). (*D*) Gel shift assays using a 760-bp dsDNA from AFV1 mixed with AFV1–132 (*Left*) or AFV1–140 (*Right*). The numbers (1–8) above the 2 × 8 lanes indicate the molar ratios of protein/dsDNA, which are: 0, 4, 10, 50, 100, 200, 500, and 1,000. (*E* and *F*) Electron Microscopy contrast views of in vitro experiments mixing Lambda DNA fragments with AFV1–132 or AFV1–140, respectively.

of uniform width and different lengths highly similar to the body of the virion (21). Plugs of ≈50 nm in length are located at both ends of the tube, and each carries three short tail fibers probably involved in host cell adsorption (19–21).

Lipothrixviruses are 0.4–2-μm-long particles, surrounded by lipid-containing envelope, and, based on different terminal structures and properties of their linear genomes, they have been classified into four genera (17, 22). To date, four protein structures from lipothrixviruses (23–26) have been determined. To complete structural information about Lipothrixviridae, and in particular to determine the structure of their MCPs, we have studied the Gamma-lipothrixvirus *Acidianus* filamentous virus 1 (AFV1). This virus infects hyperthermophilic archaea of the genus *Acidianus* living in acidic hot springs at temperature >85 °C and at pH ≈1.5. Its genome is encoded in a linear dsDNA consisting of 20,869 nucleotide base pairs and coding 40 ORFs (22). The virion is ≈9,100-Å long and 210-Å in diameter and is covered by a lipid-containing outer-shell (Fig. 1*A*). Filament both ends are capped by claw-like structures that clamp onto receptors located on host's pili during infection (Fig. 1*A*).

We report here the crystal structures of the two MCPs from AFV1. These two proteins bind DNA and form filaments when incubated with linear dsDNA. They share a similar four-helix structure. Importantly, a similar fold has been recently identified in the C-terminal domain of the coat protein from the rudivirus SIRV (27). Our data thus support at the structural level the grouping between Lipothrixviridae and Rudiviridae previously proposed at the genomic level (18) and could define a viral lineage based on the common fold of the major coat proteins from filamentous archaeal viruses. We discuss possible relationships between these MCP structures and those of eukaryotic filamentous viruses with RNA genomes and propose a topological model of the AFV1 filament derived from our structural and biochemical data.

Results

Two MCPs Form the AFV1 Core. Only two proteins from the AFV1 virion could be identified by SDS gels and a trypsin peptide mass fingerprint (Fig. 1*B*), AFV1–132 (15 kDa, pI = 9.4) and AFV1–140 (15.7 kDa, pI = 5.4), which likely form the core of the filamentous virion. The proteins share 31% sequence identity but have no significant sequence similarity with any other proteins in public sequence databases. The first 50 residues of these proteins share 46% sequence identity and are rich in basic residues, while the sequence identity of the last 40 residues decreases to 14%. Furthermore, the last helix of AFV1–140 displays a nonambiguous amphiphilic character (Fig. 1*C*).

We expressed both proteins in *Escherichia coli* (see *Materials and Methods*). Significant shifts were observed in electrophoretic mobility shift assays (EMSA) in which genomic 760-bp dsDNA was incubated with AFV1–132 or AFV1–140. This assay demonstrates that both MCPs are DNA-binding proteins that likely interact with the viral genome (Fig. 1*D*). Furthermore, in vitro experiments mixing Lambda DNA fragments with AFV1–132 or AFV1–140 gave rise to long filaments, as observed by electron microscopy (Fig. 1 *E* and *F*). These filaments seem rather flexible, as manifested by straight and bent segments. The variation in both filaments' appearance in terms of thickness might be due to unspecific DNA-protein interactions as expected for a whole genome packaging proteins. Although these filaments are not directly relevant to the virion structure, they

Table 1. Data collection and refinement statistics

	AFV1–132, S-SAD	AFV1–140, Se-SAD
Data Collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions a, b, c (Å)	28.35, 47.57, 60.08	58.16, 81.15, 119.90
Wavelength (Å)	1.54	0.979
No. of images ($\Delta\varphi = 1°$)	2000	170
Resolution (Å)	25.6–1.95 (2.06–1.95)	40–2.3 (2.42–2.3)
Unique reflections	6349 (902)	25779 (3569)
Redundancy	76 (76)	7.4 (5.2)
Completness (%)	100 (100)	99.3 (96)
$<I>/<\sigma I>$	136 (89)	14.9 (3.7)
R_{sym}%	3.7 (6.6)	10 (36)
Wilson B-factor (Å^2)	11.2	29.4
Refinement		
Resolution (Å)	18–1.95 (1.97–1.95)	15–2.3 (2.36–2.3)
Unique reflections	5710 (388)	24567 (1651)
Completness (%)	99.7 (96.4)	99.3 (93.7)
R/R_{free}	0.158/0.198 (0.157/0.190)	0.18/0.25 (0.21/0.31)
No. atoms protein/water	765/106	4397/243
B-factors (Å^2) protein/water	9.95/19.93	21.7/23
R.m.s deviations		
Bond lengths(Å)	0.013	0.012
Bond angles (°)	1.064	1.346

One crystal was used for AFV1–132 and AFV1–140 datasets. Highest-resolution shell is shown in parentheses.

indicate that both AFV1–132 and AFV1–140 may participate in nucleocapsid assembly.

Crystal Structure of the AFV1–132 Protein. Large orthorhombic crystals (250 × 31 × 15 μm) containing one cleaved monomer of AFV1–132 protein per asymmetric unit [Vm = 1.3 Å^3 Da^{-1}; (28)] were obtained after 1 week in crystallization plates (Table 1). We collected two highly redundant datasets at the Cu $K\alpha$ wavelength (λ = 1.54 Å) and solved the structure by Sulfur-SAD using an anomalous signal from three methionines (Met-74, Met-89, and Met-98), one cysteine (Cys-62), and one solvent-bound chloride ion. The structural model was refined at a resolution of 1.95 Å with R and R_{free} values of 0.158 and 0.198, respectively (Table 1), and counts for 100% in the most favored regions of the Ramachandran chart (29). The electron density map obtained for AFV1–132 contains residues between Tyr-51 and the C-terminal Tyr-132. Tight packing of the monomers and mass spectrometry measurements confirmed that AFV1–132 was cleaved during crystallization. Unexpectedly, ΔN_AFV1–132 and the C-terminal region (residues 52–134) of the basic major coat protein of the rudivirus SIRV (SIRV-134) were found to have an identical four-helix bundle fold (r.m.s deviation ≈1.4 Å for all Cα atoms) despite a very low 13% sequence identity detected only using structure-based alignment (Fig. 2*A* and *B*). We were unable to produce crystals of full-length AFV1–132. These helical domains are stabilized by an extensive network of hydrophobic interactions in the protein core. Long helices ranging from 15 (α1 and α4) to 24 residues (α2) in length are connected with short loops of approximately three residues. The two longest helices, α2 and α3, are almost parallel with each other, whereas the two shortest, α1 and α4, are slightly rotated relative to each other and are twisted at a large angle (≈45°) relative to α2 (Fig. 2*A*). The C terminus of SIRV-134 exhibits a supplementary extended stretch of eight residues, in contrast with ΔN_AFV1–132 that ends with an α-helix. A Dali search for similar structures did not return any significant hits (PDB: 2ot3, Z-score: 7.7 to a subregion of the RAB5 GDP/GTP exchange factor). Therefore, the helical domains of SIRV-134 and AFV1–132 should be considered independently folded structures. Finally, both proteins have a large positive region probably involved in DNA-binding (Fig. 2*C*).

An important characteristic shared by AFV1–132 and SIRV-134 is the disorder of their N-termini. In SIRV-134, this disorder has been demonstrated directly by NMR (27), while in AFV1–132, we deduced the disordered character of the N terminus due to its rapid cleavage both in solution and during crystallization. Cleavage of AFV1–132 also occurs at the exact place where disorder begins in SIRV-134.

Crystal Structure of the AFV1–140 Protein. Thin orthorhombic crystals of selenomethionine-labeled AFV1–140, containing four molecules per asymmetric unit (Vm = 2.21 Å^3 Da^{-1}), grew in a few days in crystallization plates. The structure was solved by SAD at the Se K-edge (λ = 0.979 Å) and was refined at a resolution of 2.3 Å with R and Rfree values of 0.18 and 0.25, respectively. The final model accounted for 96.9 and 3.1% of residues being in the most favored and generously allowed regions of the Ramachandran chart, respectively (29). Native AFV1–140 crystallized under the same conditions as selenomethionine-labeled AFV1–140 but diffracted X-rays at a low resolution. Therefore, we used the Se-SAD structure of AFV1–140 for the rest of the description.

AFV1–140 encompasses five α-helices and has an elongated shape with overall dimensions of 82.5 Å × 23.5 Å × 26 Å (Fig. 3*A*). The first 14 residues (predicted as a positively charged helix) are not visible in the electron density map. At the N terminus, the first helix α'1 (15–24) is connected to the long 53 Å helix α'2 (32–67) via a seven-residue flexible region. ΔN_AFV1–132 and the C-terminal domain of AFV1–140 have very close structures (r.m.s deviation of 1.54 Å for all Cα atoms; Fig. 3*A*) despite their low sequence identity (23%) and their different hydrophilic nature. Even though some N-terminal residues are missing, electrostatic surface properties clearly identify the first helix as the prominent positive area (Fig. 3*B*). AFV1–140 crystals are made up of one bone-like tetramer (Fig. 3*C*) per asymmetric unit. This structure can be described as a dimer of dimers. Two monomers (A and C) associate with each other head-to-tail through a 1,398-Å^2 area

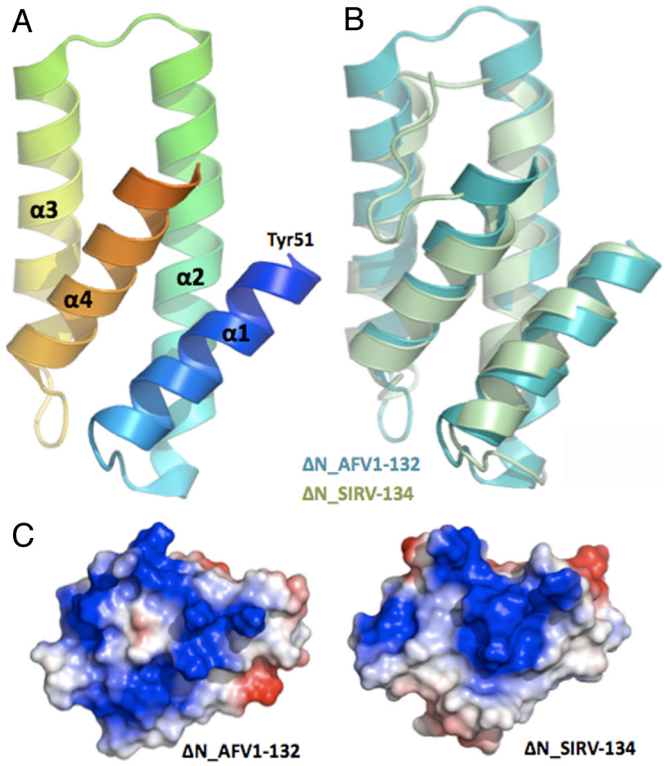

Fig. 2. Crystal structure of ΔN_AFV1–132. (*A*) Ribbon representation of ΔN_AFV1–132 crystal structure with a rainbow color gradient from the N terminus (blue) to the C terminus (red). Helices are sequentially labeled α1 to α4. Tyr-51 is the first residue seen in the electron density map. (*B*) ΔN_AFV1–132 crystal structure (cyan) superimposed onto ΔN_SIRV-134 (residues 52–134) crystal structure (light green) with a r.m.s deviation of 1.4 Å on the Cα backbone. (*C* and *D*) Electrostatic surfaces of ΔN_AFV1–132 and ΔN_SIRV-134, respectively, in the same orientation. Figs. were created with Pymol, http://pymol.sourceforge.net/.

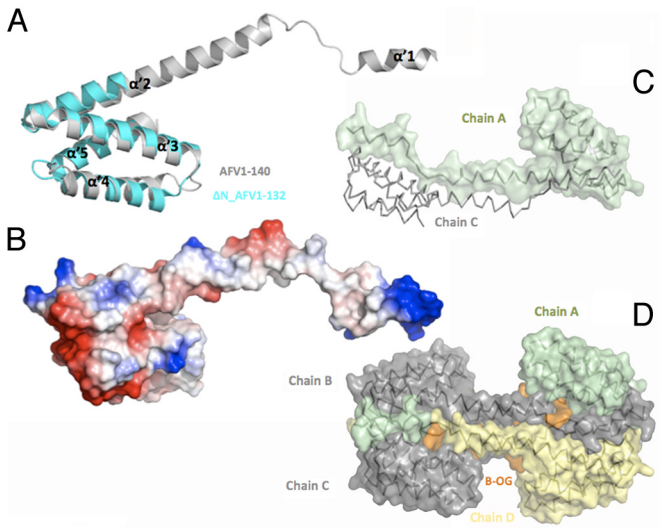

Fig. 3. Crystal structure of AFV1–140. (*A*) AFV1–140 crystal structure (gray) superimposed onto ΔN_AFV1–132 (cyan) with a r.m.s deviation of 1.54 Å on the Cα backbone. Helices are sequentially labeled α′1 to α′5. (*B*) Electrostatic surface in the same orientation. (*C*) Head-to-tail dimer in which one chain is shown as a gray ribbon and the other is represented as light green surface. (*D*) Tetramer of the asymmetric unit is composed of a dimer of dimers. Detergent molecules (β-OG) are represented with orange surfaces. Figs. were created with Pymol.

of extensive van der Waals interactions, a network of hydrogen bonds, and one salt bridge that locks the α'1 helix (Lys-19-N^{ζ}) to the C-terminal domain (Asp-73-O^{δ}) at each extremity. The same dimer BD stacks onto the other such that the extended structures connecting α'1 and α'2 helices in each monomer cross each other and hug the terminal domain (Fig. 3*D*). This assemblage is stabilized by hydrophobic interactions involving eight detergent molecules of *n*-octyl-β-D-glucopyranoside (β-OG), which are essential for purification and crystallization. Four of these molecules are buried at the C-terminal domains (191 $Å^2$ buried surface area for each), and four are buried along the α'2 helix each of domain that is involved in the buried surface area of 135.8 $Å^2$, 238.8 $Å^2$, and 53.6 $Å^2$ with monomers A, B, and β-OG_143C, respectively (Fig. 3*D*). In contrast to AFV1–132, a large part of the N terminus of AFV1–140 is ordered, with the exception of the first 14 residues. Indeed, the packing of the AFV1–140 may contribute to the stability of the N terminus. This arrangement of dimer of dimers packs each N terminus domain against the three others in the crystal packing. This arrangement also forms well-defined detergent pockets. It should be noticed that AFV1–140 alone is monomeric in solution at low concentration (<0.1 mg/mL, the upper concentration limit in the absence of detergent). When β-OG is added, the protein can be concentrated up to 5 mg/mL. However, analysis of this solution using MALLS/UV/RI (see *Materials and Methods*) yields a broad unsymmetrical peak, revealing the presence of several species in solution, possibly from monomer to tetramers. Indeed, crystallization conditions may have displaced the equilibrium toward tetramers. Such associations are probably not directly relevant to biological oligomeric states, since positively charged helices are not accessible and in opposite directions. Nevertheless, as it is always the case for viral capsid proteins, AFV1–140 as well as AFV1–132 must likely be self-associated in some way to build the virion architecture.

Discussion

We have shown here that the two MCPs of the lipothrixvirus AFV1, AFV1–132 and AFV1–140, share a common fold that is also present in the CP of the rudivirus SIRV, SIRV-134. The discovery of homology between proteins without significant similarities at the sequence level (those of the lipothrixvirus AFV1 and rudiviruses) once more exemplifies the fact that tertiary structure is more conserved than sequences (30). Importantly, our finding supports at the structural level the grouping between Lipothrixviridae and Rudiviridae into a single order, previously proposed at the genomic level, *Ligamenvirales* (18). It is widely accepted that the capsid or CPs of viruses are part of the «self» viral proteins coded by hallmark genes and that are vertically inherited (6). With this in mind and because the structure of MCPs from filamentous archaeal viruses are similar, we propose that it might be used to define a lineage of dsDNA viruses: the filamentous archaeal viruses lineage.

It should be important to know if the MCPs of *Ligamenvirales* are evolutionary related to those of viruses producing filamentous virions and infecting bacteria or eukarya or if the linear structure of these virions is an example of convergent evolution between unrelated viruses (31). To date, we know only the structures of the filamentous ssRNA Tobamovirus tobacco mosaic virus [TMV; (7)] and the ssDNA Inovirus phage fd (32). Interestingly, the structure of the CP from TMV, although mainly made up of a four-helix bundle, has a different fold compared to the MCPs of AFV1 (Fig. S1).

Biochemical and structural analysis of the MCPs of the lipothrixvirus AFV1 also sheds light on the structural organization of its virion. We found that both AFV1–132 and

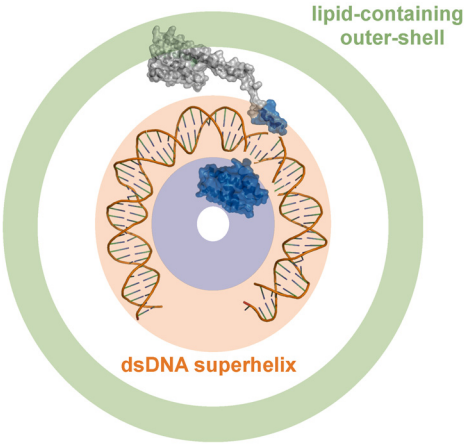

Fig. 4. Topological model of AFV1. The basic AFV1–132 patch is represented by the blue surface in the center. The dsDNA backbone (PDB :1EQZ) is orange and AFV1–140 is represented by the gray surface. AFV1–140 contacts both the lipid-containing outer-shell (in green) and the genomic DNA. In this simplified model, only one copy of each MCP has been roughly placed relatively to the genomic DNA (orange ring) and to the outer lipidic shell (green ring). A complete model would consist of a section filled by both MCPs.

AFV1–140 bind linear dsDNA and form filaments in vitro. This suggests that both proteins contribute to the formation of a filamentous nucleocapsid. The virions of filamentous ssRNA viruses such as TMV are composed of helical nucleocapsids in which the genome, wound into a superhelix, is covered by a protein shell made of several copies of one coat protein. Because the stiff-rod shaped rudivirus virion is morphologically comparable to TMV, it could be similarly formed by a helical wrapping of the CP around the nucleic-acids filament. In agreement with this hypothesis, cryo-electron microscopy pictures of native and SDS-stripped AFV1 display zipper-like or brush-like structures that are characteristics of two-dimensional projections at different angles of a dsDNA superhelix that would span the entire AFV1 core. The flexible lipothrixvirus AFV1 with two MCPs might have a more complex structure. Indeed, although their C-termini share the same globular structure, the two MCPs drastically differ in terms of physico-chemical properties: the AFV1–132 C terminus is very positively charged (pI = 9.5) as SIRV-134, and as expected for DNA-binding proteins, while the AFV1–140 C terminus exhibits a marked hydrophobic character, which is localized in its C-terminal helix, and is significantly soluble only in the presence of detergent. Because both proteins bind DNA, we suggest a model in which the DNA wraps around AFV1–132 whereas AFV1–140 N terminus might be attached to the DNA while its globular lipophilic C-terminal domain would be more exposed at the surface and could become part of this envelope forming a lipido-proteic layer (Fig. 4). One can suggest a sequence of evolutionary events in which the unique CP of a "simpler" non-enveloped virion, such as those of rudiviruses, have been duplicated and evolved so as to facilitate interactions with a hydrophobic envelope, producing the more complex virions of lipothrixviruses.

Materials and Methods

Production of AFV1–132 and AFV1–140. AFV1–132 and AFV1–140 were cloned into the Gateway™ destination vector pETG-20A (Dr. Arie Geerlof, EMBL Hamburg). The final construct encoded a thioredoxin fusion protein containing an N-terminal hexahistidine tag followed by a TEV protease recognition site upstream of the protein of interest. Fusion proteins were overexpressed in the *E. coli* Rosetta(DE3) pLysS strain (Novagen) grown in Terrific Broth (Gibco) at 37 °C for 4 h and at 25 °C overnight for AFV1–132 and AFV1–140, respectively (33). Overexpressed proteins were purified as follows: (*i*) using nickel affinity chromatography (His-Trap 5 mL column, GE Healthcare) with a step-gradient of imidazole; (*ii*) performing TEV protease cleavage; (*iii*) recovering cleaved proteins in the flowthrough of a second nickel affinity chromatography; and (*iv*) performing a preparative size-exclusion chromatography on a Superdex 200 HR26/60 in Tris 10 mM, pH 8, and NaCl 300 mM and on a Superdex 200 HR10/30 in Tris 10 mM, pH 8, NaCl 300 mM, and 30 mM *n*-octyl-β-D-glucopyranoside (Anagrade, Anatrace) for AFV1–132 and AFV1–140, respectively. Samples of AFV1–140 with *n*-octyl-β-D-glucopyranoside were analyzed using analytical size exclusion chromatography (SEC) on a HPLC system (Waters) with on-line multi-angle laser light scattering, absorbance and refractive index detectors (MALLS/UV/RI) (Wyatt Technology). The SEC was performed on a Shodex KW-403 4F column with Tris 10 mM, pH 8, NaCl 300 mM, and 30 mM *n*-octyl-β-D-glucopyranoside as the eluent. Selenomethionine-labeled AFV1–140 was prepared following standard procedures in the minimum medium M9 by blocking the methionine biosynthesis pathway. Expression, purification, and characterization of the SeMet-labeled AFV1–140 were carried out using the same protocols as for native protein.

Electron Microscopy of the MCPs and DNA. AFV1–140 and DNA filaments were generated using a solution containing 13 mM $CaCl_2$, 13 mM $MgCl_2$, 1.3 mM NaN_3, and 83 ng/mL Lambda DNA cut with *Eco*RI and *Hin*dIII, and 1 mg/mL AFV1–140. The solution was adjusted to a pH of 3 with 5% H_2SO_4 and incubated at 78 °C for 7 days, and filaments were negatively stained with 2% uranyl acetate. Filaments of AFV1–132 and DNA were generated mixing 50 ng/μL Lambda DNA cut with *Eco*RI and *Hin*dIII and 3 mg/mL AFV1–132. Filaments were then incubated at 60 °C in a 10 mM Hepes buffer with a pH of 7 that contained 12.8 mM $CaCl_2$, 12.8 mM $MgCl_2$, and 1.28 mM NaN_3. Filaments were negatively stained with 1.5% phosphotungstic acid and were observed after 4 days of incubation on a 200 kV electron microscope at magnification, ×20,000.

Electrophoretic Mobility Shift Assay. PCR-generated 760-bp dsDNA fragments from the AFV1 genome were independently incubated with AFV1–132 and AFV1–140 in TBE buffer (90 mM Tris-borate, pH 8, and 2 mM EDTA) supplemented with 40% glycerol at increasing dsDNA/protein molar ratios of 1/0, 1/4, 1/10, 1/50, 1/100, 1/200, 1/500, and 1/1,000. After 15 min incubation at 55 °C and then overnight incubation at room temperature, reaction products were visualized by ethidium bromide staining after electrophoresis for 8 h at 1 V/cm through 1.0% agarose gels run in TBE buffer.

Crystal Structures. Crystallization of AFV1–132 was performed at 20 °C using the hanging-drop vapor-diffusion technique. Diffraction quality crystals (250 × 31 × 15 μm) were obtained after 1 week by mixing 1 μL protein at 12 mg/mL with 1 μL reservoir solution containing 12% PEG 8000 and 25 mM Na_2HPO_4/KH_2PO_4 with a final pH of 6. MALDI-TOF performed on crystals returned a molecular mass for the cleaved protein of ≈9,000 Da. After cryo-cooling in reservoir solution supplemented with 10% glycerol (vol/vol), a SAD experiment was performed in-house at the Cu $K\alpha$ wavelength (λ = 1.54 Å) using a Bruker Axis Microstar rotating anode. Two SAD datasets were collected from a single crystal at different χ angles using a MAR345dtb image plate detector. Data were integrated with MOSFLM and scaled with SCALA (34). The PHENIX software suite was used to solve the structure (35). HYSS readily identified five peaks arising from three methionines, one cysteine, and one solvent-bound chloride ion. Initial phases were calculated with PHASER and were improved using RESOLVE (36). The structural model was automatically built with Arp/wArp 6.0 (37) and then refined at 1.95 Å using REFMAC5 (38).Crystallization of selenomethionine-labeled AFV1–140 was performed at 20 °C using the sitting-drop vapor-diffusion technique at 20 °C and a nanodrop-dispensing robot (Cartesian Inc.) in 96-well Greiner crystallization plates. Crystals grew by mixing 100 nL protein at 4.3 mg/mL with 100 nL of 17% PEG 550 MME and 0.1 M sodium cacodylate with a final pH of 5. A SAD experiment was performed on beamline ID14–4 (ESRF) at the Se K-edge (λ = 0.979 Å). One SAD dataset was collected at 100 K from a single crystal using an ADSC Q4 detector. Data were processed using MOSFLM/SCALA (34). Selenium substructure (16 sites) was solved using SHELXD and initial phases were calculated after density modification using SHELXE (39). Solvent flattening, noncrystallographic symmetry (NCS) averaging and partial model building were performed using RESOLVE (36, 40). This model was manually completed using COOT (41) and was refined at a resolution of 2.3 Å using REFMAC5 (38) and by applying three TLS groups (Translation/Libration/Screw rotation) (42) per monomer with R and Rfree values of 0.18 and 0.25, respectively (Table 1). The final model accounted for 96.9% and 3.1% of residues being in most

favored and generously allowed regions of the Ramachandran chart, respectively (29). Native AFV1–140 crystallized in the same condition as the selenomethionine-labeled protein. One dataset was collected at the ESRF beamline ID14–2 using an ADSC Q4 detector. Native structure was solved by molecular replacement.

ACKNOWLEDGMENTS. We thank Prof. Michael G. Rossmann for productive discussions and Dr. Arie Geerlof for kindly providing us with the Gateway plasmid pETG-20A for the His-Trx fusion. This work was supported by the Agence Nationale pour la Recherche (ANR France) Grant ANR-05-BLAN-0089 and by Marseille-Nice Genopole.

1. Forterre P (2006) The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res* 117:5–16.
2. Koonin EV, Senkevich TG, Dolja VV (2006) The ancient Virus World and evolution of cells. *Biol Direct* 1:29.
3. Raoult D, Forterre P (2008) Redefining viruses: Lessons from Mimivirus. *Nat Rev Microbiol* 6:315–319.
4. Moreira D, Lopez-Garcia P (2009) Ten reasons to exclude viruses from the tree of life. *Nat Rev Microbiol* 7:306–311.
5. Forterre P, Prangishvili D (2009) The origin of viruses. *Res Microbiol* 160:466–472.
6. Krupovic M, Bamford DH (2008) Virus evolution: How far does the double beta-barrel viral lineage extend? *Nat Rev Microbiol* 6):941–948.
7. Sachse C, et al. (2007) High-resolution electron microscopy of helical specimens: A fresh look at tobacco mosaic virus. *J Mol Biol* 371:812–835.
8. Tars K, Fridborg K, Bundule M, Liljas L (2000) The three-dimensional structure of bacteriophage PP7 from *Pseudomonas aeruginosa* at 3.7-A resolution. *Virology* 272:331–337.
9. Lescar J, et al. (2001) The fusion glycoprotein shell of Semliki Forest virus: An icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* 105:137–148.
10. Pletnev SV, et al. (2001) Locations of carbohydrate sites on alphavirus glycoproteins show that E1 forms an icosahedral scaffold. *Cell* 105:127–136.
11. Bamford DH, Grimes JM, Stuart DI (2005) What does structure tell us about virus evolution? *Curr Opin Struct Biol* 15:655–663.
12. Nandhagopal N, et al. (2002) The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. *Proc Natl Acad Sci USA* 99:14758–14763.
13. Rice G, et al. (2004) The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proc Natl Acad Sci USA* 101:7716–7720.
14. Khayat R, et al. (2005) Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses. *Proc Natl Acad Sci USA* 102:18944–18949.
15. Benson SD, Bamford JK, Bamford DH, Burnett RM (1999) Viral evolution revealed by bacteriophage PRD1 and human adenovirus coat protein structures. *Cell* 98:825–833.
16. Baker ML, Jiang W, Rixon FJ, Chiu W (2005) Common ancestry of herpesviruses and tailed DNA bacteriophages. *J Virol* 79:14967–14970.
17. Prangishvili D, Garrett RA, Koonin EV (2006) Evolutionary genomics of archaeal viruses: Unique viral genomes in the third domain of life. *Virus Res* 117:52–67.
18. Prangishvili D, Forterre P, Garrett RA (2006) Viruses of the Archaea: A unifying view. *Nat Rev Microbiol* 4:837–848.
19. Prangishvili D, et al. (1999) A novel virus family, the Rudiviridae: Structure, virus-host interactions and genome variability of the sulfolobus viruses SIRV1 and SIRV2. *Genetics* 152:1387–1396.
20. Vestergaard G, et al. (2005) A novel rudivirus, *ARV1*, of the hyperthermophilic archaeal genus *Acidianus*. *Virology* 336:83–92.
21. Vestergaard G, et al. (2008) Structure of the acidianus filamentous virus 3 and comparative genomics of related archaeal lipothrixviruses. *J Virol* 82:371–381.
22. Bettstetter M, Peng X, Garrett RA, Prangishvili D (2003) AFV1, a novel virus infecting hyperthermophilic archaea of the genus *Acidianus*. *Virology* 315:68–79.
23. Keller J, et al. (2007) Crystal structure of AFV3–109, a highly conserved protein from crenarchaeal viruses. *Virol J* 4:12.
24. Keller J, et al. (2009) Crystal structure of AFV1–102, a protein from the *Acidianus* filamentous virus 1. *Protein Sci* 18:845–849.
25. Goulet A, et al. (2009) The thermo- and acido-stable ORF-99 from the archaeal virus AFV1. *Protein Sci* 18:1316–1320.
26. Goulet A, et al. (2009) The crystal structure of ORF14 from *Sulfolobus islandicus* filamentous virus. *Proteins* 76:1020–1022.
27. Szymczyna BR, et al. (2009) Synergy of NMR, computation, and X-ray crystallography for structural biology. *Structure* 17:499–507.
28. Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* 33:499–501.
29. Laskowski RA, Moss DS, Thornton JM (1993) Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* 231:1049–1067.
30. Rossmann MG, Moras D, Olsen KW (1974) Chemical and biological evolution of nucleotide-binding protein. *Nature* 250:194–199.
31. Ackermann HW (2007) 5500 Phages examined in the electron microscope. *Arch Virol* 152:227–243.
32. Wang YA, et al. (2006) The structure of a filamentous bacteriophage. *J Mol Biol* 361:209–215.
33. Vincentelli R, et al. (2003) Medium-scale structural genomics: Strategies for protein expression and crystallization. *Acc Chem Res* 36:165–172.
34. CCP4 CCPN (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760–766.
35. Adams PD, et al. (2002) PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* 58:1948–1954.
36. Terwilliger TC (2003) Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr D Biol Crystallogr* 59:38–44.
37. Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6:458–463.
38. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240–255.
39. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64:112–122.
40. Terwilliger TC (2000) Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr* 56:965–972.
41. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
42. Winn MD, Murshudov GN, Papiz MZ (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol* 374:300–321.